

UNITED STATES DISTRICT COURT  
DISTRICT OF NEVADA

Cung Le, Nathan Quarry, and Jon Fitch, on  
behalf of themselves and all others similarly  
situated

Plaintiffs,

v

Zuffa LLC, d/b/a Ultimate Fighting  
Championship and UFC,

Defendants

SECOND DECLARATION OF  
CHARLES KELLNER

I, Charles Kellner, hereby declare:

1. I testified previously in this matter and provided in that declaration my qualifications and a statement of the scope of my inquiry. (Declaration of Charles Kellner, dated February 19, 2016 Document Number 221).

2. Since the date of that declaration, I attended the Status Conference on February 23, 2016 and I have reviewed the transcript of that conference. I attended telephone conferences among counsel for the parties on March 25, March 30, April 22, and April 25. I have reviewed correspondence and attachments from the parties that set forth their positions with respect to the scope of discovery herein.

3. The following updates my prior testimony combined with additional analysis.

**LACK OF PROGRESS IN NEGOTIATING SEARCHES**

4. In my last declaration, I wrote that **use of search terms can sometimes be a blunt instrument**. Search terms can be overly broad or not broad enough to capture the information

1 requested. A difficulty in this case is that of searching for evidence of particular kinds of  
2 behavior. Use of idiom can make it difficult to identify the terms to be used. Plaintiffs want  
3 search terms that yield information about *behavior*. Defendant wants to rely on searches  
4 designed only to yield documents about *transactions*.  
5

6 5. The financial services industry provides some instruction about how to search for  
7 particular kinds of behavior. For example, the regulators have the advantage of making sure  
8 broker-dealers' emails are preserved for a period of years according to SEC and FINRA  
9 requirements. FINRA tells us that a combination of random and targeted sampling, recording the  
10 use of (with continuing updates to) the lexicon, and targeted full review of communications are  
11 the best-practices backbone for monitoring compliance.  
12

13 Lexicon-Based Reviews of Electronic Correspondence –Members using lexicon  
14 based reviews (those based on sensitive words or phrases, the presence of which  
15 may signal problematic communications) of correspondence should utilize an  
16 appropriate lexicon, take reasonable security measures to keep the list confidential  
17 and periodically evaluate the efficacy of the lexicon. Members must make  
18 informed decisions regarding how best to utilize the surveillance tools they have  
19 chosen.... Members should also consider regular periodic reviews of the lexicon  
20 system to determine whether any changes/updates are necessary, such as adding  
21 or deleting phrases and/or words....As discussed more fully below, if a member  
22 does not have confidence in the effectiveness of its lexicon system, a  
23 supplemental random review of electronic communications should be considered.  
24

25 Members should consider targeted concentrated reviews of employees' emails  
26 when warranted (*e.g.*, when concerns are raised in connection with regulatory  
27 examination findings, internal audits, customer complaints or regulatory  
28 inquiries). When assessing the effectiveness of a lexicon-based system, members  
should consider the following features: (a) A meaningful list of phrases and/or  
words (including industry "jargon") based on the size of the member, its type of  
business, its customer base and its location (including any branch offices that may  
require the inclusion of certain foreign language components). The lexicon system  
should be comprehensive enough to yield a meaningful sample of "flagged"

1 communications. (b) Ability to add and delete phrases and words on an ongoing  
2 basis.

3 Combination of Lexicon and Random Review of Electronic Correspondence –  
4 Given the strengths and weaknesses of any single review tool, members should  
5 consider complementary review techniques....[A] registered representative  
6 determined to circumvent a lexicon system may be able to do so by simply  
avoiding the use of words likely to “trigger” the system.<sup>1</sup>

7 6. In the era before or in the absence of advanced analytics or TAR, then, the regulators  
8 are telling us to have complete control of the collection, to continuously evaluate and update the  
9 search terms, to be on guard for changes in idioms already scrutinized, and to initiate random  
10 samples, targeted samples and targeted full reads. That approach is a fine ideal for regulators, but  
11 not available in civil litigation, and nothing to which Defendants have even remotely agreed.  
12

13 7. Defendant has resisted providing any more search hits than those already ordered by  
14 the Court.<sup>2</sup> The most progress Defendant has offered in the past two months is a new search  
15 report proffered on April 25, 2016. It contains 296 search phrases almost all of which have  
16 proximity limiters of “within 5.” More than half of the offered search phrases yield zero hits. All  
17 told they yield fewer than 16,000 documents out of a putatively searchable set of more than  
18 940,000.<sup>3</sup> Moreover, the hits were calculated before an industry standard deduplication process  
19  
20

---

21  
22 <sup>1</sup> Regulatory Notice 07-59 “Supervision of Electronic Communications: FINRA Provides  
23 Guidance Regarding the Review and Supervision of Electronic Communications”, December  
24 2007. <https://www.finra.org/sites/default/files/NoticeDocument/p037553.pdf> last researched  
April 26, 2016

25 <sup>2</sup> Email letter from M. Lynch to M. Dell’Angelo dated April 22, 2016. “Here are the potentially  
26 disputed topics that Zuffa may wish to address in the Joint Status Report: 1. Whether, to what  
27 extent, and how Zuffa will be required to engage in additional review of documents beyond that  
already ordered or agreed to.”

28 <sup>3</sup> Email correspondence by M. Lynch to Counsel dated April 25, 2016 with attached Exhibit  
P(1)-Zuffa Additional Search Term Hit Report.xlsx

1 was applied, meaning that likely fewer than 16,000 documents were actually captured by these  
2 new terms.

3 8. This is not progress, but rather further indication of Defendant's entrenchment. The  
4 parties are not getting anywhere in discussing or negotiating for the right search terms.

5 Defendant appears willing to search for and produce documents about transactions, but resists  
6 approaches that would produce documents about behavior. This most recent offer of search terms  
7 is evocative of the discussion between the parties prior to the last Status Conference.  
8

9 9. By contrast, Plaintiffs have been pressing the case for "winnowing" out of the  
10 collection categories of documents that Defendants can remove from the Remainder set without  
11 review. At last count, that winnowing and other factors identified below have had the impact of  
12 reducing the remainder that Plaintiff wants searched by more than 300,000 documents.  
13

14 10. The impact of that reduction means that Plaintiffs proposed search terms are now  
15 considerably more precise in locating responsive documents as previously. The search hits on the  
16 remainder, including documents with families, by Defendants' own search report, is 603,431,  
17 This number includes Plaintiff's requested search hits for Twitter and Facebook searches that  
18 increase the count by a substantial number due to Defendants' inability to screen out their email  
19 signature lines from being included in the searches.<sup>4</sup>  
20  
21

22  
23 <sup>4</sup> Although the search term hit report for Plaintiffs' search terms show more than 140,000  
24 document hits with families for each of the terms "Facebook" and "Twitter," the number of  
25 documents hit by these terms that are not also hit by others of Plaintiffs' search terms is unclear.  
26 In late March, Defendant, at Plaintiffs' request, ran hit reports on the non-duplicated set of 1.4  
27 million Email Documents (the set discussed at the last status conference), to test the effects of  
28 those two terms. Despite the non-duplicated hit counts for each term exceeding 300,000  
(without families), when both terms were excluded from search, fewer than 50,000 documents  
were removed from the reviewable population returned by Plaintiffs' search terms (only 4.2% of  
that population). If, these figures hold true for the current set, notwithstanding that each term

11. But there is significantly more information to convey that has developed over the past two months that serve to demonstrate Defendant's entrenchment.

#### **ROUTINE DIFFICULTY IN AGREEMENTS ON FILTERING**

12. In my prior declaration I made several points that **Defendant is not helping its own cause**. Plaintiffs have been working on multiple simultaneous suggestions for how Defendant can potentially "winnow" down the collection or otherwise by identifying and removing clearly non-responsive documents from searches. Defendant's execution of these suggestions has not evidenced a willingness to work toward progress.

- a. Defendant continues not to assist actively in contributing to discriminating keywords or in otherwise filtering the collection.
- b. Selection of domain names that can be included or excluded has stalled.
- c. Impact of documents to exclude, such as certain "form documents" Plaintiffs have proposed to filter and exclude have yet to be quantified.
- d. Removal of Outlook contact cards from the search set continues to be a challenge.
- e. Agreement on search terms to isolate and exclude privileged materials from the Legal Custodians' files have only recently been finalized.
- f. Defendant has not provided information on the impact several of its meager proposals would have on the set.
- g. Defendant has been unable or unwilling to quantify the review impact of its own email threading on the rate or burden of review.

---

generates more than 140,000 hits, removing them would remove fewer than 30,000 documents from the reviewable population in Plaintiffs' Search Term Set.

1 h. Defendant has refused to provide metrics on the rate of responsiveness for  
2 documents that hit on its search terms.  
3

4  
5 **OVERCOUNTING AND LACK OF DE-DUPLICATION,**  
6 **AGAINST FINANCIAL SELF-INTEREST**

7 13. One of the more disturbing discoveries over the past two months is how Defendant  
8 de-duplicated its collection and how that method of de-duplication served significantly to inflate  
9 Defendant's document counts.

10 14. Just before the prior Status Conference, Plaintiffs received a sampling from  
11 Defendant that contained a set of unsearchable PDF images. Plaintiffs' counsel identified  
12 documents that were in all respects identical and duplicative except for the seconds counted in  
13 the date stamp. That discovery led to Plaintiff's investigation of how the Defendant de-  
14 duplicated and counted its documents in its representations and reports.  
15

16 15. On April 6, 2016, Defendant stated that its de-duplication process removed only  
17 2,066 "email documents" (totaling 11,575 documents across the network email and hard drive  
18 documents) out of a universe of approximately 2,000,000 documents. Deduplication at a rate of  
19 0.01% is shockingly low and should have caused the Defendant or its vendor to investigate the  
20 deduplication method applied and explore appropriate means of deduplication to reduce its  
21 claimed burden of review.  
22

23 16. Plaintiffs' counsel requested technical specifications about de-duplication on March  
24 11, 2016 in light of their discovery. Defendant responded on March 21, 2016 that it used "the  
25 most conservative form of de-duplication which ensures that only identical documents, based on  
26 their metadata fields, will be excluded." "This means that documents whose substantive text may  
27 be identical may not be treated as duplicates because their respective metadata may vary." *Id.*  
28

1 Plaintiffs followed up on March 23, 2016 requesting the “exact protocol” used by Defendant’s  
 2 vendor to deduplicate the Email Documents and the number of documents that were removed by  
 3 application of that process. Defendants’ counsel replied on March 28, again stating that de-  
 4 duplication was performed using a “conservative” protocol, and vaguely described a strict hash  
 5 comparison of emails. Plaintiffs requested that Zuffa arrange a conference call that included a  
 6 senior member of Zuffa’s vendor’s operations group to describe the protocol performed. That  
 7 call took place on March 31.

8  
 9 17. On the March 31 call, in which I participated, Plaintiffs learned that Defendant’s  
 10 method of de-duplication would not de-duplicate the following, even if the senders, recipients,  
 11 subject line, text and attachments are all identical:

- 12 a. Sent and received items of the same email in the same email box.
- 13 b. Sent and received copies of the same email in two or more different mailboxes.
- 14 c. cc-ed items of the same email
- 15 d. Sent and received items of the same email, if the sender is also a recipient on a  
 16 distribution list.

17  
 18 18. Defendants’ method of deduplication only de-duplicated the same email if it was  
 19 found in the same email box of the same custodian or in a copy of that custodian’s email box.

20  
 21 19. Defendant used a method of “hashing” emails. Hashing calculates a value based on  
 22 every letter and space in a document, in order, and yields a unique string of characters. There are  
 23 different algorithms used for hashing, either “SHA” or “MD5” but they all work essentially the  
 24 same, and none are likely to yield a false duplicate in more than a billion tries. An “MD5” hash  
 25 value of “Ckellner” is 59d4bf20dee9efc0565e041a32415a3b. The same algorithm applied to  
 26  
 27  
 28

1 “ckellner” is 9757a67fcf998757d28b755125c0e1c0, so even changing the case of a letter yields a  
2 completely different unique value.

3 20. Most eDiscovery software and service providers use hashing to compare “loose files”  
4 for duplicates. These are average MS Office documents and PDFs found in millions of desktops,  
5 laptops, servers, mobile devices and as attachments to emails.  
6

7 21. EDiscovery software and service providers do not purely hash email messages, as  
8 Defendants did, to identify duplicates. The reason is that emails with exact content have different  
9 time stamps, if only by one second, for the sent and received items, wherever they land. Even if  
10 sent across the hall, the differences in sent time and received time create differences that will be  
11 detected by hashing and, under the method applied by the Defendant, not identify the documents  
12 as duplicates.  
13

14 22. Email messages also have different email “routing” information, which is the log of  
15 the email’s travels through the email systems and the Internet. An email received outside the  
16 enterprise has a long and rich routing history that makes its hash different than its identical  
17 counterpart in the sender’s email box.  
18

19 23. To further illustrate, Lexis-Nexis, the maker of one of the most common eDiscovery  
20 processing tools, compares email metadata fields and compares hashes only for attachments.  
21 <http://help.lexisnexis.com/litigation/ac/law/law6.6/deduping.htm> Similarly, Craig Ball  
22 addressed the concept of “[h]ashing the ESI as a file (i.e., a defined block of data) containing the  
23 ESI using the same hash algorithm (e.g., MD5 or SHA1) and comparing the resulting hash value  
24 for each file. If they match, the files hold the same data. This tends not to work for e-mail  
25 messages exported as files because, when an e-mail message is stored as a file, messages that we  
26 regard as identical in common parlance (such as identical message bodies sent to multiple  
27  
28



recipients) are not identical in terms of their byte content. The differences tend to reflect either variations in transmission seen in the message header data (the messages having traversed different paths to reach different recipients) or variations in time (the same message containing embedded time data when exported to single message storage formats as discussed above with respect to the .MSG format).” A Ball In Your Court, July 4, 2012.

24. Plaintiffs and Defendant agreed on global de-duplication, meaning that they had an expectation that they would de-duplicate identical items among senders and recipients.

***Defendants’ chosen method of de-duplication did not de-duplicate across custodians.*** Because a sender and recipient would have different routing information in the email, the method chosen by Defendants to de-duplicate would not identify as duplicates most messages that are identical but for the fact that one was stored in the sender’s Sent Items and the other in the recipient’s Inbox. The method of de-duplication selected by Defendants defeated the purpose and intent of global de-duplication.

25. The Sedona Conference Principles Cooperation Guidance for Litigators and In-House Counsel states that

The parties should discuss how to minimize the review and production of duplicate materials without an undue risk of losing unique documents. This discussion can involve both collection issues (collecting duplicative data from different sources or custodians) and also the electronic de-duplication methodology to be applied to the data after it is collected. The production of duplicate information can become quite expensive. Both parties should be motivated to agree on a de-duplication protocol. Proper de-duplication will reduce costs for both sides.

Parties should agree on de-duplication methods. For example, they might de-duplicate documents only within each custodian (so that an email will appear in each user’s mailbox from which it was collected), or they might de-duplicate across the entire database of collected documents (so that an email appears only

once regardless of how many persons received it so long as tracking information regarding the email going to other custodians is provided). Usually when de-duplication is done across the entire database, the vendor is able to access information showing each of the mailboxes that received the email.<sup>5</sup>

26. The standards used by the eDiscovery software and services industry to de-duplicate emails are designed according to these Sedona principles. They are designed for predictability in agreements among counsel, and they are designed to reduce counts and to save money. Those methods of de-duplication do, in fact, hash attachments, message text, and email metadata about sender, recipient, cc, and subject, but they specifically and purposefully ignore time stamps and email routing information.

27. As far as cost is concerned,

If ediscovery were a small part of litigation and duplicate consolidation had an imperceptibly small impact on ediscovery, the whole debate might be dismissed under the rationale of de minimis non curat lex. However, the cost of ediscovery in general, and the cost of relevance and privilege reviews in particular, have been a major concern for years. There are no excuses for “not getting it” when it comes to ediscovery. Lawyers who bill hundreds of dollars an hour are implicitly promising a certain level of competence that would include the basic notion of consolidating duplicates. The survey shows that respondents had been offering duplicate consolidation for eight to 10 years unless they hadn’t been in business that long. When lawyers and the courts were first grappling with electronically stored information (ESI), maybe there was an excuse for not having thought this issue through, but those days are gone.<sup>6</sup>

---

<sup>5</sup> The Sedona Conference “Cooperation Guidance for Litigators & In-House Counsel,” March 2011 (emphasis added). This, in my opinion, is the very simple and bullet-point guide of what to do in eDiscovery, without a lot of technical jargon.

<sup>6</sup> Patrick Oot, Joe Howie & Anne Kershaw of the Ediscovery Institute republished “Ethics and EDiscovery Review” in ACC Docket, Vol 28 Issue 1 (Jan-Feb 2010) were referring in this passage to Rule 1.1, Competence.

1           28. Defendant's vendor knew that Zuffa's email deduplication method was not designed  
2 to remove most email duplicates.

3           29. For example, R. Losey, in eDiscovery Blog, March 19, 2012, estimates that as high as  
4 30-40% deduplication in some cases if done globally. Lange and Nimsger indicate in the treatise  
5 "Ediscovery- What Every Lawyer Should Know" that deduplication can be as high as 90% in  
6 some cases but an average of 30-40%. Kroll Ontrack Pulse Benchmarks December 2014  
7 indicates that more than 60% of projects choose case-level rather than custodian level de-  
8 duplication.  
9

10           30. Defendant's counsel and expert are experienced in de-duplication techniques.  
11 Defendant's counsel and I discussed the prospect of "near duplicate detection" along with email  
12 threading as early as February 2016, and I was told it was not available with this vendor.  
13 Defendant's expert was an employee of a well-known eDiscovery software vendor, a member of  
14 a large law firm's eDATA practice group, and had "unfettered access to the service provider and  
15 they have provided complete answers to my questions about the capabilities of the software..."  
16

17           31. Despite the industry standard and the stipulation of the parties to de-duplicate across  
18 custodians, Defendant applied its "conservative" method. Two weeks after Plaintiffs requested  
19 the number of documents that were de-duplicated out of the set using this method, Defendant  
20 finally provided the information. By letter on April 6, 2016, Defendant informed Plaintiffs that  
21 only 2,066 emails and attachments were de-duplicated by its process, approximately 0.01% of  
22 the corpus.  
23

24           32. Such a meager reduction from de-duplication is abnormally low by industry  
25 standards. My own experience is at least 25% de-duplication and higher when I see a relatively  
26 small group of witnesses who tend to work closely together.  
27  
28

1           33. Upon application of an industry standard de-duplication process which Plaintiffs  
2 proposed with my assistance, Defendant determined that 223,191 documents could be removed  
3 from the set as duplicates, more than 108 times more than Defendant's process.  
4

5           34. Although this proportion of documents (less than 11% of the corpus) is low by  
6 industry standards, I understand that the time periods for the custodial email collections are not  
7 precisely aligned. By this I mean that some custodians have emails from as early as 2010 while  
8 others are limited to a time period beginning in early 2014. I believe this accounts for all or part  
9 of the difference between the actual proportion deduplicated and the typical experience.  
10

#### 11           **APPARENT INDIFFERENCE TO AVAILABILITY OF TECHNOLOGY**

12           35. In my prior declaration, I proposed that perhaps Defendant's technology may not be  
13 up to the task. Before and at the last Status Conference, Plaintiffs raised the issues that  
14 Defendant's eDiscovery technology may be under-powered or outdated. Defendant countered  
15 that its vendor's technology is proprietary, but that its functionality is nevertheless mainstream.  
16 Now my opinion is that regardless of the technology available to Defendant through its current  
17 vendor, Defendant is not taking advantage of it. For example:  
18  
19

- 20           a. Defendant's email collection has a sizeable number of emails that contain  
21 "marketing footers" or signature lines. These contain graphics and also links to  
22 Defendant's websites, and links to Defendant-sponsored websites, among them,  
23 Facebook pages and Twitter feeds. The presence of these website links in the  
24 footers complicates Plaintiff's requests to search for references to Facebook,  
25 Twitter, and others that are critical to Plaintiffs' discovery requests about  
26 sponsors. Many mainstream eDiscovery applications can employ analytics to  
27  
28

discount “repetitive content” so that these phrases can be searched without hitting on the repetitive content in the signature lines. It is the same technology that litigants use to discount “privilege boilerplate” in email signature lines that interfere with searching for privileged documents. Defendant represents that it does not have the technology to avoid this repetitive content, but the website of its vendor, Rational Enterprise, indicates otherwise.<sup>7</sup>

- b. Running and re-running search reports requested by Plaintiffs have often taken a week or more between iterations, contrary to the experience of Boies Schiller and its vendor, Rational Enterprise, as represented on the vendor’s website.<sup>8</sup>

36. These instances combined with the indifference to the impact of Defendant’s “conservative” method of de-duplication, against its financial interest and against industry standard, incline me toward the opinion that Defendant is more concerned about the prospect of producing the emails about behavioral content which Plaintiffs seek in the Remainder Set than it is about the cost of reviewing those emails.

#### **RATE OF REVIEW MITIGATES AGAINST INCREMENTAL APPROACH**

---

<sup>7</sup> “No more are attorneys weeding through thousands of irrelevant documents because they happened to respond to poorly crafted search terms; and no more are emails mistakenly identified as privileged or work product simply because those words appear in an email signature.” <http://www.rationalenterprise.com/resources/files/the-rational-enterprise-ediscovery-playbook.pdf> (last viewed on April 26, 2016).

<sup>8</sup> Defendant’s vendor Rational Enterprises website contains a testimonial from Defendant’s own counsel Boies Schiller stating, in a recent case with 16 million documents, among other things, as follows “Over 35,000 individual searches were run on the system and on many days more than 1,000 searches were run, many of which were run simultaneously.” (last viewed at <http://www.rationalenterprise.com/resources/case-studies/adam-r-shaw-partner> on April 25, 2016)

1 37. Defendant's first production based on the Court's prior order arrived on April 6,  
2 2016. By that date, approximately one month after the date of the court's order, Defendant had  
3 reviewed, according to Defendant's counsel on the call dated March 25, only about 50,000 out of  
4 more than 595,000 documents required to be reviewed. Notably, at that point, Defendant had not  
5 applied Plaintiffs' de-duplication protocol or all of its winnowing and two-tier search proposals  
6 which reduced the size of the reviewable population.  
7

8 38. At that rate it will be more than a year before Plaintiffs can determine whether they  
9 have received the requested and required production about Defendant's internal conduct. Further,  
10 Plaintiff's own analysis shows that of the 35,698 custodial documents produced as of April 8,  
11 2016, there are only 4,835 unique message threads.<sup>9</sup> Thus, after more than a month, Defendant  
12 has produced fewer than 5,000 email conversations. There is nothing about this "wait and see"  
13 process that will work effectively to determine whether Plaintiffs yet have the discovery they  
14 need. Even at the rate of a Status Conference every two months, with this approach we will be  
15 four to five more conferences away from progress.  
16  
17

18 39. Since the date of that production, Defendants have refused to divulge any more  
19 information about their rates of review, or how many documents reviewed yielded how many  
20 responsive documents produced. This lack of progress gives Plaintiffs no comfort or even basic  
21 information that search hits that are being reviewed are either largely responsive or not  
22 responsive, and Defendants appear indifferent to the impact. The rate of responsiveness of  
23  
24

---

25 <sup>9</sup> Because Defendant did not produce the conversation metadata, Plaintiffs relied on normalized  
26 subject line data (*i.e.*, the subject line without terms like "re:" or "Fwd:"), the message dates and  
27 address matching. This method does not separate side conversations, so the number of properly  
28 threaded emails is likely higher than the 4,835, but not substantially. Regardless, the clustering of  
even properly threaded emails (*i.e.*, clustering of the core conversation and side conversations),  
would achieve the same review facilitation benefits.

1 Defendant's search terms is a useful metric because it helps the parties and the Court to evaluate  
2 the reliability of the Defendant's search terms. However, it is only a guide..  
3

#### 4 5 **APPROACH TO RESOLUTION**

6 40. In my previous declaration I pointed to the distance between the parties in negotiating  
7 search terms. I posited that if Defendant did not want to review search hits and was concerned  
8 about too many non-responsive search hits, that it could apply a privilege screen and produce the  
9 rest under an unconditional clawback. Defendant has since stated explicitly that it was unwilling  
10 to produce without review.  
11

12 41. My opinion is that Defendant is unwilling further to negotiate on search terms and  
13 that it has no concrete proposal on producing from the portions of the collection currently not  
14 under order by the search terms previously ordered. There are now, suddenly, approximately 1.8  
15 million more documents from the 2011-and-earlier time period that were previously presumed  
16 gone.  
17

18 42. The Defendant and its counsel have engaged in a lot of activity, but that activity has  
19 not been directed necessarily toward narrowing the gap between Plaintiffs discovery requests and  
20 Defendants' professed but now unsupportable and not credible claims for proportionality. For  
21 example, in my experience, I have never seen a party use Defendant's patently ineffective  
22 method of "conservative" de-duplication prior to review, particularly for the review of a large set  
23 of documents, where it is arguing about or legitimately concerned about cost and burden.  
24

25 43. In the last Joint Status Report and expert declaration, Defendant complained about  
26 cost and burden of additional discovery. Plaintiffs stated in their documents and at the hearing  
27 that Defendant was over-counting its burden in numerous ways. When Plaintiffs complained that  
28

1 search reports with counts would each take a week, Defendant managed to speed them up. When  
2 Plaintiffs expressed concern that including the “unique hits” in search reports would take an  
3 additional week per report, Defendant managed to include them with the report.  
4

5 44. When Plaintiff complained prior to the last Status Conference that the hits for  
6 individual terms with “families” was yielding an overcount in search hits, on the morning of the  
7 conference Defendant demanded that Plaintiffs retract the statement. Hours later, Defendant then  
8 explained to the Court—what it would not explain to Plaintiffs—how it derived these “with  
9 family” hit counts. Plaintiffs evaluated that statement and informed Defendant precisely how  
10 Defendants were overcounting these hits. *See* Exhibit A (which was provided to Defendant as an  
11 attachment to Plaintiffs’ March 11, 2016 Letter). Defendant has now conceded that it had  
12 overcounted in the way Plaintiffs described. As a result, Defendant’s reports to the Court search  
13 terms like “Hold%” were returning half a million hits were dramatically overstated. In fact,  
14 based on the current hit report, in the Remainder Set, as currently formulated, Plaintiffs’ term  
15 “Hold%” returns only 26,142 documents with families.  
16  
17

18 45. I have specialized training and more than twenty-five years of experience in the  
19 impact and effect of applying key words to full text in the context of litigation discovery. I have  
20 more than eighteen years of experience in assisting litigants to manage the scope of electronic  
21 discovery and in evaluating eDiscovery procedures and associated costs. I have had, variously,  
22 leadership roles, CLE teaching experience, writing responsibilities, or active participation with  
23 the ABA Legal Technology Advisory Council, the ABA section for Employment Litigation  
24 (Professional Responsibility Committee), the Electronic Discovery Reference Model (EDRM),  
25 the Georgetown Law Center Advanced EDiscovery Institute, Massachusetts MCLE, and The  
26 Sedona Conference, among others. At D4 and elsewhere I have assisted with keyword filtering in  
27  
28



1 hundreds of cases, for both requesting and responding parties. I have testified as an expert in the  
2 past on the costs and burdens of eDiscovery and have served as neutral third party expert in  
3 several cases in United States District Courts.  
4

5 46. I do not see the parties making progress in negotiating search criteria or agreeing to a  
6 complex protocol of “full reads” in conjunction with random and targeted sampling. My opinion  
7 is that the scope of discovery, particularly with respect to discovery of email sought by Plaintiffs  
8 about Defendant’s behavior, will need to be ordered by the Court. Defendant seems more  
9 concerned about the prospect of producing this material than about the cost of its review. In my  
10 experience I have never seen a producing party, as here, actually inflate the review set. If the  
11 Defendant was truly concerned with reducing the burden of its review, it could have  
12 independently undertaken many of the proposals that Plaintiffs developed over the past several  
13 months. Importantly, as the same time, Defendant has done little to nothing to improve the  
14 precision and recall of its own search terms. Rather it has simply relied upon Plaintiffs to reduce  
15 its burden through difficult and time consuming investigatory work.  
16  
17

18 47. Plaintiffs should get the benefits of the search terms they have laboriously developed  
19 and the time and effort it has put into helping the Defendant filter its own collection. In the  
20 alternative, the Court should order a full review and production, subject to a privilege screen and  
21 unconditional clawback for documents not previously produced.  
22

23 Respectfully submitted on this 29th day of April  
24

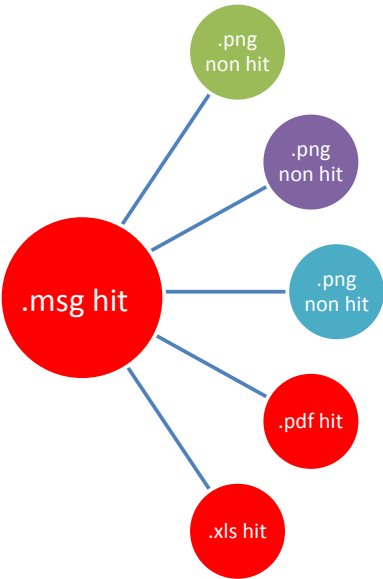
25 /s/Charles R. Kellner  
26

27 Charles R. Kellner  
28

# Exhibit A

## Representation of Number of Documents with Family

### 1 Email with Attachment Hits



### Zuffa’s Hit Report for that Family

Document	Hit	With Family
Email Hit	1	6
.xls Hit	1	6
.pdf Hit	1	6
Hit Report Entry	3	18

### De-Duped Hit Report for that Family

Document	Hit	With Family
Email Hit	1	6
.xls Hit	1	6
.pdf Hit	1	6
Hit Report Entry	3	6